Global (Q)SAR models on substrates for human Cytochrome P450 3A4

Ringsted, Tine; Nikolov, Nikolai Georgiev; Jensen, Gunde Egeskov; Wedebye, Eva Bay; Niemelä, Jay Russell

Publication date:
2010

Document Version
Early version, also known as pre-print

DTU Food
National Food Institute

Global (Q)SAR models on substrates for human Cytochrome P450 3A4

Tine Ringsted, Nikolai Nikolov, Gunde Egeskov Jensen, Eva B. Wedebye and Jay Niemelä
National Food Institute, Technical University of Denmark
E-mail: {tiri, jarn, nign and gunje}@food .dtu.dk

The aim of this study was to develop QSAR predictions for identifying CYP3A4 substrates and non-substrates in the EINECS chemicals (European INventory of Existing Commercial chemical Substances).

Introduction

The CYP3A subfamily plays a dominant role by the elimination of around 35% of prescribed drugs. It has a broad catalytic selectivity and metabolises among others drugs, steroids, carcinogens and pesticides. The substrates range in size from small molecules e.g. acetaminophen (151 g/mol) to relatively big molecules such as Cyclosporin A (1203 g/mol).

Besides helping with the excretion of exogenous compounds CYP3A4 has also shown to transform protoxicans into their toxic structure. An example is Aflatoxin B_1 (AFB_1) which needs to be transformed before it can function as a hepatic carcinogen. CYP3A4 metabolise AFB_1 and produces both genotoxic and non-genotoxic metabolites.

Materials and Methods

Modelling methodology

MultiCASE Groups the chemicals by structure (2-10 interconnected non-hydrogen atoms) and performs multiple linear regression on these groups. A modulator (e.g. activating fragments, inactivating fragments, $logK_{ow}$, molecular orbital energies) is used in each linear regression as independent variable and the activity as the dependent variable.

Leadscope The chemicals are classified by structural features into categories e.g. functional groups, heterocycles, pharmacophores and $logK_{ow}$. The important structural features are then used in a partial least squares (PLS) model. The PLS model reduces the number of structural features and these features are used as the independent variables in a logistic regression.

SciQSAR The linear parametric discriminant analysis was used. 86 physical-chemical descriptors were used as independent variables. The variables describe e.g. electron accessibility, the skeletal branching and $logK_{ow}$.

Data

•The training set consists of experimental human data on 863 chemicals[1,2,3,4] . 503 positive and 360 negative as CYP3A4 substrates.

•40,0374 EINECS chemicals (European INventory of Existing Commercial chemical Substances).

Results

Mean and range results for cross-validation

	Sensitivity (correct pos)	Specificity (correct neg)	Concordance (correct total)
MultiCASE[1]	**68.1**	**52.1**	**61.5**
	67.4-68.9	51.5-53.0	60.8-62.3
Leadscope[1]	**78.9**	**61.6**	**71.7**
	77.9-79.5	59.7-63.3	71.1-72.8
SciQSAR[2]	**75.7**	**48.3**	**64.3**

[1]: Leave 50% out cross-validation were performed 3 times

[2]: LOO cross-validation was performed

Predictions for EINECS chemicals

	In domain [%]	In domain and predicted to be a substrate [%]
MultiCASE	**49**	**33**
Leadscope	**34**	**25**
SciQSAR	**100**	**66**

Discussion

25% - 66% of the EINECS chemicals were predicted as CYP3A4 substrates. SciQSAR predicts the double amount of CYP3A4 substrates in the EINECS chemicals compared to MultiCASE and Leadscope. The actual number of CYP3A4 substrates in the EINECS chemicals is therefore still difficult to approximate. However, since 35% of prescribed drugs are metabolised by CYP3A4, the predictions by MultiCASE and Leadscope which say that 25-33% of EINECS chemicals are metabolized by CYP3A4 can be considered to be realistic. The number of CYP3A4 substrates is most likely overestimated in the SciQSAR model.

Conclusion

The models produced by Leadscope and MultiCASE can be used to predict CYP3A4 substrates in untested chemicals and identify potential toxicity concerns due to exposure to environmental chemicals.

References:

1. Yap,C.W., Chen,Y.Z. Prediction of cytochrome p450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines, Journal of Chemical Information and Modeling. 2005, 45, 4, 982-992.

2. Rendic S., Summary of information on human CYP enzymes: Human P450 metabolism data, Drug Metabolism Reviews. 2002, 34, 83-448.

3. Sheridan R.P., Korzekwa K.R., Torres R.A. and Walker M.J., Empirical Regioselectivity Models for Human Cytochromes P450 3A4, 2D6, and 2C9, Journal of Medicinal Chemistry. 2007, 50, 14, 3173-3184.

4. U.S. National Library of Medicine, Daily Med Current Medication Information, Accessed March-April 2009, Available at http://dailymed.nlm.nih.gov/dailymed/about.cfm